ISSN: 2993-8643
Research Article
Engineering: Open Access

Transformer CS-EEG: A Transformer-Based Deep Learning Framework for Compressed Sensing of EEG Signals

Zhiying Xu[1,2*] and Huotao Gao[2]

[1]School of Electronic Information, Wuhan University, Wuhan, China

[2]School of Information Engineering, Wuhan Business University, Wuhan, China

[*]Corresponding Author
Zhiying Xu, School of Electronic Information, Wuhan University, Wuhan, China.

Submitted: 2025, Sep 08; **Accepted:** 2025, Oct 01; **Published:** 2025, Oct 06

Citation: Xu, Z., Gao, H. (2025). Transformer CS-EEG: A Transformer-Based Deep Learning Framework for Compressed Sensing of EEG Signals. *Eng OA, 3*(10), 01-07.

Abstract
Electroencephalogram (EEG) signals present significant challenges for efficient acquisition and processing due to their high dimensionality and complex temporal-spatial patterns. This paper presents Transformer CS-EEG, a novel Transformer-based deep learning framework for compressed sensing of EEG signals. Our approach leverages the self-attention mechanism to effectively capture long-range dependencies and complex spatial-temporal correlations in EEG data. We introduce three key innovations: (1) a specialized EEG sampling encoder with adaptive learning of optimal sampling patterns and channel-wise attention mechanisms, (2) frequency-aware multi-head self-attention specifically tailored for EEG characteristics that targets different neurophysiological frequency bands, and (3) a multiscale reconstruction decoder that progressively recovers signal details through hierarchical upsampling. Extensive experiments on three public EEG datasets (CHB-MIT for seizure detection, BCI Competition IV for motor imagery, and DEAP for emotion recognition) demonstrate that our approach consistently outperforms state-of-the-art methods across various compression ratios, achieving up to 18% lower reconstruction error and 2.4 dB higher signal-to-noise ratio at 10× compression. Furthermore, our method preserves essential neurophysiological information, maintaining over 96% of the original accuracy in downstream EEG analysis tasks. The proposed framework offers a promising solution for efficient EEG data acquisition in resource-constrained environments while ensuring high-quality signal reconstruction for accurate clinical interpretation and brain-computer interface applications.

Keywords: Transformer CS-EEG, Network-driven Deep Learning Architecture, Compressed Sensing, Electroencephalogram Signals

1. Introduction

Electroencephalogram (EEG) signals, which measure electrical activity of the brain through scalp electrodes, are fundamental tools in modern neuroscience and clinical neurophysiology. They are widely used in brain-computer interfaces (BCIs) for assistive technology and neurorehabilitation, neurological disorder diagnosis including epilepsy and sleep disorders, and cognitive state monitoring for attention and workload assessment. The non-invasive nature and high temporal resolution of EEG make it invaluable for understanding brain dynamics and developing neural interfaces.

Traditional EEG acquisition systems require high sampling rates (typically 250-1000 Hz) to capture neural dynamics accurately, resulting in massive data volumes that pose significant challenges. A standard clinical EEG recording with 64 channels at 500 Hz generates approximately 15 MB of data per minute, leading to storage requirements exceeding 20 GB for 24-hour monitoring. This data deluge creates bottlenecks in storage, transmission, and real-time processing, particularly problematic in resource-constrained environments such as wearable devices, remote monitoring systems, and point-of-care diagnostics.

Compressed sensing (CS) has emerged as a promising paradigm to address these challenges by enabling signal acquisition at sub-Nyquist rates while ensuring accurate reconstruction through exploitation of signal sparsity [1]. The classical CS theory relies on two fundamental assumptions: sparsity of the signal in some transform domain and incoherence between the sampling and sparsity domains. Traditional CS approaches for EEG typically employ random sensing matrices (Gaussian or Bernoulli) for signal acquisition, followed by iterative optimization algorithms such as Basis Pursuit, Orthogonal Matching Pursuit (OMP), or Approximate Message Passing (AMP) for signal reconstruction [2].

However, these traditional methods face several critical limitations when applied to EEG signals. First, EEG signals are non-stationary with time-varying spectral characteristics, making it challenging to identify a single optimal sparsifying transform. Second, the iterative reconstruction algorithms suffer from high computational complexity ($O(N^3)$ for many algorithms), making them unsuitable for real-time applications. Third, these methods often fail to capture the complex spatial-temporal correlations inherent in multi-channel EEG recordings, leading to suboptimal reconstruction quality, especially at high compression ratios.

Recent advances in deep learning have revolutionized the CS landscape by enabling data-driven approaches that learn both sampling and reconstruction processes directly from data. Convolutional neural networks (CNNs) have been widely adopted for CS applications, with notable examples including ReconNet, which pioneered end-to-end learning for CS reconstruction, and ISTA-Net, which unrolls iterative algorithms into trainable neural networks [3,4]. These methods have demonstrated superior performance compared to traditional approaches by implicitly learning complex signal structures without requiring explicit sparsity assumptions.

Despite their success, CNN-based approaches are inherently limited by their local receptive fields, which restrict their ability to model long-range dependencies. This limitation is particularly problematic for EEG signals, where neural oscillations exhibit complex temporal dynamics spanning multiple time scales and spatial correlations across distant brain regions. The sliding window operation of CNNs processes signals locally, potentially missing global patterns crucial for accurate reconstruction.

Transformer networks, originally proposed for natural language processing, have emerged as powerful architectures for sequence modeling through their self-attention mechanism [5]. Unlike CNNs, Transformers can directly compute relationships between any pair of positions in a sequence, regardless of their distance. This global receptive field makes them par ticularly suitable for capturing the long-range dependencies and complex correlations in EEG signals. The multi-head attention mechanism allows the model to attend to different representation subspaces simultaneously, potentially capturing various aspects of EEG dynamics.

In this paper, we propose Transformer CS-EEG, a novel Transformer-based deep learning framework specifically de signed for compressed sensing of EEG signals. Our approach addresses the limitations of existing methods through several key innovations:

1) End-to-end Learning Framework: We develop a unified architecture that jointly optimizes both sampling and reconstruction processes, leveraging the self-attention mechanism to capture long-range temporal dependencies and cross-channel spatial correlations that are characteristic of EEG signals.

2) Adaptive Sampling Encoder: We design a specialized sampling module with learnable convolutional filters that adaptively optimize sampling patterns for different EEG channels and frequency bands, coupled with a channel attention mechanism that dynamically weights the importance of different electrodes based on their information content.

3) Frequency-Aware Attention: We propose a novel multi-head self-attention mechanism that incorporates frequency-specific filtering matrices, allowing different attention heads to specialize in capturing patterns from distinct EEG frequency bands (delta, theta, alpha, beta, gamma) that carry different neurophysiological significance.

4) Multi-Scale Reconstruction: We develop a hierarchical decoder that progressively reconstructs signals from coarse to fine scales through a combination of trans posed convolutions and dilated convolutions, effectively capturing both slow oscillations and fast transients in EEG signals.

We conduct extensive experiments on three diverse public EEG datasets to validate our approach: CHB-MIT for seizure detection, BCI Competition IV Dataset 2a for motor imagery classification, and DEAP for emotion recognition. Our results demonstrate that Transformer CS-EEG consistently out performs state-of-the-art methods across various compression ratios, achieving significant improvements in reconstruction quality while preserving task-relevant neurophysiological in formation.

2. Related Work

2.1. Compressed Sensing for EEG Signals

Traditional CS approaches for EEG signals have focused on identifying appropriate sparsifying transforms. Discrete wavelet transform (DWT) has been widely used due to EEG's multi-scale nature, while discrete cosine transform (DCT) and empirical mode decomposition (EMD) have also shown promise. However, these fixed transforms often fail to adapt to the non-stationary characteristics of EEG signals. Several works have explored structured sensing matrices specifically designed for EEG, including Kronecker-based approaches that leverage spatial-temporal correlations and data-driven optimization of sensing matrices. Despite these advances, traditional methods remain computationally intensive and struggle with high compression ratios.

2.2. Deep Learning for Compressed Sensing

The integration of deep learning with CS has led to significant breakthroughs. CNN-based architectures have dominated this space, with networks learning to map compressed measurements directly to reconstructed signals. Recent approaches have incorporated residual connections, recurrent layers for temporal modeling, and dual-domain processing combining time and frequency information. However, these methods are limited by local receptive fields and struggle to capture the global dependencies crucial for EEG signal reconstruction.

3. Methodology

3.1. Problem Formulation

In the compressed sensing framework for EEG signals, we consider a multi-channel EEG recording $X \in \mathbb{R}^{C\times N}$, where C represents the number of electrode channels and N is the number of temporal samples. The compressed measurements $Y \in \mathbb{R}^{C\times M}$ are obtained through a sampling process:

$$Y = \Phi X + n \quad (1)$$

where $\Phi \in \mathbb{R}^{M\times N}$ is the sampling matrix with $M < N$, and n represents measurement noise. The compression ratio is defined as CF = N/M indicating the degree of dimensionality reduction.

Traditional CS approaches solve an optimization problem with sparsity constraints, assuming the signal is sparse in some transform domain Ψ:

$$\min_X \|\Psi X\|_1 \quad \text{subject to} \quad \|Y - \Phi X\|_2 \le \varepsilon \quad (2)$$

In contrast, our deep learning approach learns both the optimal sampling strategy and reconstruction function directly from data:

$$\hat{X} = f_\theta\left(\Phi_\phi(X)\right) \quad (3)$$

where Φϕ represents a learnable sampling operation with parameters ϕ, and f_θ is the reconstruction network with parameters θ.

3.2. Transformer CS-EEG Architecture

Our proposed Transformer CS-EEG architecture consists of three main components that work synergistically to achieve high-quality EEG reconstruction from compressed measurements.

1) Adaptive Sampling Encoder: The sampling encoder implements an adaptive, learnable compression strategy that goes beyond traditional random sampling matrices. For an input EEG segment $X \in \mathbb{R}^{C\times N}$, we first apply channel-wise 1D convolution with learnable filters:

$$Y_{conv} = \text{Conv1D}(X; W_{conv}) \quad (4)$$

where $W_{conv} \in \mathbb{R}^{1\times N\times M}$ are learnable convolutional filters that span the entire signal length to capture temporal patterns. This operation effectively implements a learnable sampling matrix that can adapt to the statistical properties of EEG signals during training.

To address the varying information content across EEG channels, we incorporate a channel attention mechanism:

$$A_{ch} = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot \text{AvgPool}(Y_{conv}) + b_1) + b_2) \quad (5)$$

$$Y = A_{ch} \odot Y_{conv} \quad (6)$$

where $A_{ch} \in \mathbb{R}^{C\times 1}$ represents channel-wise attention weights, ⊙ denotes element-wise multiplication, and $\boldsymbol{\sigma}$ is the sigmoid activation function.

To preserve temporal ordering information crucial for sequence modeling, we add sinusoidal position encoding:

$$P_{pos}[t, 2i] = \sin(t/10000^{2i/d}) \ , \ P_{pos}[t, 2i+1] = \cos(t/10000^{2i/d}) \quad (7)$$

The final input to the Transformer module is obtained by projecting the compressed measurements to a higher dimensional embedding space:

$$Z_{in} = W_{emb} \cdot Y + P_{pos} \quad (8)$$

where $Z_{in} \in \mathbb{R}^{C\times M\times d}$, $W_{emb} \in \mathbb{R}^{M\times d}$, and d=128 is the embedding dimension\

2) Transformer-based Latent Representation Module: The core of our framework processes the embedded compressed measurements through L = 6 Transformer encoder layers. Each layer performs the following operations:

$$Z'_l = Z_{l-1} + \text{MHSA}(\text{LN}(Z_{l-1})) \quad (9)$$

$$Z''_l = Z'_l + \text{CCA}(\text{LN}(Z'_l)) \quad (10)$$

$$Z_l = Z''_l + FFN(LN(Z''_l)) \quad (11)$$

where LN denotes layer normalization, MHSA is multi-head self-attention, CCA is cross-channel attention, and FFN is the feed-forward network.

Frequency-Aware Multi-Head Self-Attention: EEG signals exhibit distinct oscillatory patterns across different frequency bands, each associated with specific brain states and cognitive processes. To capture these frequency-specific characteristics, we propose a novel attention mechanism. For each attention head $h \in \{1, 2, ..., H\}$ where H = 8:

$$Q_h = Z_{l-1} W_h^Q \quad (12)$$

$$K_h^f = Z_{l-1} W_h^K \cdot F_h \quad (13)$$

$$V_h^f = Z_{l-1} W_h^V \cdot F_h \quad (14)$$

where $F_h \in \mathbb{R}^{d_h\times d_h}$ is a learnable frequency filtering matrix. We initialize each Fh as a band-pass filter targeting different EEG frequency ranges: delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13

Hz), beta (13-30 Hz), and gamma (30-45 Hz).

The attention scores and outputs are then computed as:

$$A_h = \text{softmax}\left(\frac{Q_h \bullet (K_h^f)^T}{\sqrt{d_h}}\right) \quad (15)$$

$$H_h = A_h V_h^f \quad (16)$$

where $A_h \in \mathbb{R}^{C\times M\times M}$ represents the attention weights and $H_h \in \mathbb{R}^{C\times M\times d_h}$ is the output of attention head h.The scaling factor $\sqrt{d_h}$ prevents the dot products from growing too large in magnitude, which could push the softmax function into regions with extremely small gradients.
The outputs from all heads are concatenated and projected:

$$\text{MHSA}(Z_{l-1}) = \text{Concat}(H_1, H_2, \ldots, H_H) W^0 \quad (21)$$

Cross-Channel Attention: To model spatial correlations across EEG channels arising from volume conduction and functional connectivity:

$$\text{CCA}(Z) = \text{softmax}\left(\frac{Z W^{CQ} (Z W^{CK})^T}{\sqrt{d}}\right) Z W^{CV} \quad (22)$$

Position-wise Feed-Forward Network: Each position is transformed independently:

$$\text{FFN}(Z) = W2\ \text{GELU}(W1 Z + b1) + b2 \quad (24)$$

where GELU is the Gaussian Error Linear Unit activation, and the feed-forward dimension is $d_{ff} = 4d = 512$.
3) Multi-Scale Reconstruction Decoder: The decoder progressively reconstructs the original signal resolution through a hierarchical architecture. Starting from the Transformer output $Z_{out} \in \mathbb{R}^{C\times M\times d}$ we first project to the decoder embedding space:

$$R_0 = W_{proj} Z_{out} + b_{proj} \quad (20)$$

We then apply a series of upsampling stages, each doubling the temporal resolution:

$$R_i^{temp} = \text{ConvTranspose1D}(R'_{i-1}, W_{up}^i, b_{up}^i) \quad (21)$$

$$R_i^{dil} = \text{Conv1D}(R'_{i-1}, W_{dil}^i, b_{dil}^i, \text{dilation} = 2^{i-1}) \quad (22)$$

$$R_i^{fus} = \text{Upsample}(R'_{i-1}) \odot \sigma(W_{gate}^i R_i^{temp} + b_{gate}^i) \quad (23)$$

$$R'_i = R_i^{temp} + R_i^{dil} + R_i^{fus} \quad (24)$$

This process continues for stages until the original temporal resolution is recovered. The final reconstruction is obtained through:

$$\hat{X} = \text{Conv1D}(R'_{\log_2(N/M)}, W_{out}, b_{out}) \quad (25)$$

3.3. Loss Function

We employ a composite loss function that captures different aspects of signal fidelity:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{MSE} + \lambda_2 \mathcal{L}_{freq} + \lambda_3 \mathcal{L}_{struct} \quad (26)$$

The mean squared error loss ensures point-wise accuracy:

$$\mathcal{L}_{MSE} = \frac{1}{CN}\sum_{c=1}^{C}\sum_{n=1}^{N}\left(X_{c,n} - \hat{X}_{c,n}\right)^2 \quad (27)$$

The frequency domain loss preserves spectral characteristics:

$$\mathcal{L}_{freq} = \frac{1}{CN}\sum_{c=1}^{C}\left\|FFT(X_c) - FFT(\hat{X}_c)\right\|_2^2 \quad (28)$$

The structural similarity loss maintains morphological features:

$$\mathcal{L}_{struct} = 1 - \text{SSIM}(X, \hat{X}) \quad (29)$$

where $\lambda 1 = 1.0$, $\lambda 2 = 0.5$, and $\lambda 3 = 0.3$ are weighting coefficients determined through validation experiments.

4. Experimental Setup

4.1. Datasets

We evaluate our proposed framework on three publicly available EEG datasets:

CHB-MIT Scalp EEG Database: This dataset contains EEG recordings from 23 patients with intractable seizures. The data were collected at a sampling rate of 256 Hz using the international 10-20 electrode placement system with 23 channels.

BCI Competition IV Dataset 2a: This dataset comprises EEG recordings from 9 subjects performing four different motor imagery tasks. The data were collected using 22 electrodes at a sampling rate of 250 Hz.

DEAP Dataset: This dataset contains EEG recordings from 32 participants watching music videos. The EEG signals were recorded from 32 channels at a sampling rate of 128 Hz.

All signals are preprocessed with a 0.5-45 Hz bandpass filter to remove baseline drift and high-frequency noise. We segment continuous recordings into non-overlapping 2-second windows, resulting in N = 512 samples for CHB-MIT and BCI datasets, and N = 256 for DEAP. Data are normalized to zero mean and unit variance per channel. We use a 70/15/15 split for training, validation, and testing, ensuring subject-wise separation to prevent data leakage.

B. Implementation Details

Our model is implemented in PyTorch with the following specifications:

- Architecture: Embedding dimension d = 128, L = 6 Transformer layers, H = 8 attention heads, feed-forward dimension dff =

512, decoder embedding dr = 64

- Training: Adam optimizer with learning rate 1 × 10−4, batch size 64, cosine annealing schedule with warm restarts, 200 epochs
- Regularization: Dropout rate 0.1 after each sub-layer, gradient clipping at norm 1.0
- Hardware: NVIDIA Tesla V100 GPU with 16GB memory, training time approximately 7-8 hours per dataset

C. Evaluation Metrics

We employ comprehensive metrics to assess reconstruction quality:

Normalized Mean Square Error (NMSE): NMSE $=\|X-\hat{X}\|_F^2 / \|X\|_F^2$, measuring overall reconstruction accuracy.

Signal-to-Noise Ratio (SNR): SNR $=10 \log_{10} \|X\|_F^2 / \|X-\hat{X}\|_F^2$ in dB, quantifying signal quality.

Structural Similarity Index (SSIM): Assesses preservation of structural information through luminance, contrast, and structure comparison.

Correlation Coefficient (CC): Measures linear correlation between original and reconstructed signals.

Power Spectral Density Error: Evaluates frequency do main reconstruction quality across EEG bands.

D. Baseline Methods

We compare against representative traditional and deep learning methods:

- **Traditional CS:** OMP [2] (greedy algorithm), AMP (message passing with denoising)
- **Deep Learning:** ReconNet [3] (pioneering CNN approach), ISTA-Net [4] (algorithm unrolling), DU DORNet (dual-domain processing), CS-ConvRNN (recurrent architecture for EEG)

5. Results and Discussion

5.1. Reconstruction Performance

Table I presents comprehensive reconstruction performance across three datasets at different compression ratios. Transformer CS-EEG consistently achieves superior performance compared to all baseline methods, with improvements becoming more pronounced at higher compression ratios.

On the CHB-MIT dataset, which contains complex seizure patterns and artifacts, our method achieves 0.078 NMSE at CR=4, representing a 16.1% improvement over CS-ConvRNN. At the challenging CR=10, Transformer CS-EEG maintains 0.192 NMSE and 15.95 dB SNR, demonstrating 18.3% lower error and 2.43 dB higher SNR than the best baseline. The high correlation coefficient (0.901) at CR=10 indicates excellent preservation of temporal dynamics crucial for seizure detection.

For the BCI Competition dataset, which requires preservation of subtle sensorimotor rhythms, our approach shows 18.8% lower NMSE at CR=10.The DEAP dataset, containing affective EEG patterns, shows even greater improvements with 18.7% lower NMSE, suggesting our method's effectiveness in preserving complex emotional signatures in frontal-temporal regions.

TABLE I

Reconstruction Performance Comparison Across Datasets (Mean ± Standard Deviation)

Method	CR = 4			CR = 8			CR = 10		
	NMSE	SNR (dB)	CC	NMSE	SNR (dB)	CC	NMSE	SNR (dB)	CC
CHB-MIT Scalp EEG Database									
OMP	0.163±0.019	15.09±0.89	0.917±0.015	0.285±0.029	10.87±0.93	0.848±0.022	0.368±0.038	9.14±1.05	0.802±0.030
AMP	0.158±0.017	15.37±0.85	0.920±0.014	0.273±0.027	11.24±0.91	0.853±0.021	0.352±0.036	9.53±1.01	0.809±0.028
ReconNet	0.132±0.015	16.54±0.78	0.933±0.012	0.235±0.023	12.67±0.84	0.872±0.018	0.309±0.031	10.82±0.93	0.834±0.024
ISTA-Net	0.124±0.013	17.06±0.74	0.937±0.011	0.221±0.021	13.21±0.81	0.879±0.017	0.292±0.029	11.35±0.91	0.843±0.023
DU-DORNet	0.107±0.011	18.35±0.68	0.946±0.009	0.198±0.019	14.26±0.75	0.892±0.015	0.264±0.026	12.43±0.85	0.859±0.020
CS-ConvRNN	0.093±0.008	19.49±0.63	0.954±0.007	0.175±0.016	15.37±0.69	0.905±0.012	0.235±0.023	13.52±0.79	0.876±0.018
TransformerCS-EEG	**0.078±0.006**	**21.28±0.57**	**0.967±0.005**	**0.143±0.012**	**17.65±0.61**	**0.925±0.009**	**0.192±0.019**	**15.95±0.71**	**0.901±0.014**
BCI Competition IV Dataset 2a									
OMP	0.179±0.022	14.62±0.95	0.906±0.018	0.297±0.033	10.53±0.98	0.838±0.025	0.383±0.042	8.77±1.10	0.792±0.033
AMP	0.171±0.020	14.92±0.91	0.912±0.017	0.285±0.031	10.87±0.95	0.844±0.024	0.369±0.040	9.13±1.07	0.799±0.031
ReconNet	0.145±0.017	16.03±0.83	0.925±0.014	0.246±0.027	12.29±0.89	0.865±0.020	0.321±0.035	10.43±0.99	0.825±0.027
ISTA-Net	0.136±0.015	16.54±0.79	0.931±0.013	0.232±0.025	12.84±0.86	0.871±0.019	0.304±0.033	10.97±0.96	0.834±0.026
DU-DORNet	0.118±0.013	17.73±0.74	0.941±0.011	0.209±0.022	13.81±0.80	0.885±0.017	0.277±0.030	12.03±0.91	0.849±0.023
CS-ConvRNN	0.104±0.011	18.78±0.68	0.948±0.009	0.186±0.019	14.85±0.74	0.898±0.015	0.247±0.027	13.12±0.85	0.866±0.021
TransformerCS-EEG	**0.085±0.008**	**20.53±0.62**	**0.962±0.007**	**0.151±0.015**	**17.08±0.67**	**0.919±0.012**	**0.203±0.022**	**15.48±0.77**	**0.893±0.017**
DEAP Dataset									
OMP	0.159±0.021	15.58±0.92	0.920±0.016	0.275±0.031	11.41±0.95	0.859±0.023	0.358±0.040	9.71±1.09	0.814±0.031
AMP	0.152±0.019	15.95±0.88	0.924±0.015	0.263±0.029	11.80±0.93	0.864±0.022	0.344±0.038	10.04±1.06	0.820±0.029
ReconNet	0.129±0.016	17.09±0.81	0.937±0.013	0.228±0.025	13.12±0.87	0.882±0.019	0.301±0.033	11.31±0.97	0.844±0.025
ISTA-Net	0.121±0.014	17.67±0.77	0.941±0.012	0.215±0.023	13.67±0.84	0.889±0.018	0.285±0.031	11.87±0.94	0.852±0.024
DU-DORNet	0.105±0.012	18.91±0.72	0.950±0.010	0.193±0.020	14.71±0.78	0.902±0.016	0.259±0.028	12.97±0.89	0.868±0.022
CS-ConvRNN	0.091±0.009	20.03±0.66	0.958±0.008	0.170±0.017	15.82±0.72	0.914±0.014	0.229±0.024	14.02±0.83	0.884±0.019
TransformerCS-EEG	**0.075±0.007**	**21.69±0.60**	**0.970±0.006**	**0.139±0.014**	**18.07±0.65**	**0.932±0.011**	**0.186±0.020**	**16.37±0.74**	**0.908±0.016**

TABLE II
FREQUENCY BAND PSD ERROR AT CR = 8 (CHB-MIT DATASET)

Method	Delta (0.5-4 Hz)	Theta (4-8 Hz)	Alpha (8-13 Hz)	Beta (13-30 Hz)	Gamma (30-45 Hz)
OMP	0.192±0.022	0.218±0.025	0.265±0.029	0.325±0.033	0.401±0.040
AMP	0.183±0.020	0.209±0.024	0.255±0.028	0.317±0.032	0.395±0.039
ReconNet	0.156±0.017	0.178±0.020	0.218±0.024	0.285±0.029	0.368±0.036
ISTA-Net	0.143±0.015	0.165±0.018	0.204±0.022	0.274±0.028	0.359±0.035
DU-DORNet	0.124±0.013	0.142±0.016	0.178±0.019	0.246±0.025	0.337±0.033
CS-ConvRNN	0.109±0.011	0.125±0.014	0.159±0.017	0.227±0.023	0.321±0.031
Ours	**0.087±0.009**	**0.103±0.011**	**0.130±0.014**	**0.187±0.019**	**0.278±0.027**

5.2. Frequency Band Preservation

Table II demonstrates the superior preservation of neurophysiologically relevant frequency bands, which is critical for clinical interpretation and analysis. Transformer CS-EEG achieves the lowest PSD error across all frequency bands at CR=8.

The most significant improvements are observed in the alpha band (18.2%reduction), beta band (17.6%reduction), and gamma band (13.4%reduction) compared to CS-ConvRNN. These higher frequency bands are typically more challenging to preserve due to their lower power and higher complexity. The frequency-aware attention mechanism enables different

Attention heads to specialize in specific frequency ranges, leading to better spectral fidelity across the entire EEG spectrum.

5.3. Downstream Task Performance

Table III evaluates the preservation of task-relevant information, demonstrating that reconstruction quality translates to maintained performance in practical applications.

For motor imagery classification, Transformer CS-EEG maintains 96.7%of the original accuracy, enabling reliable BCI control evenwith8×compression.The preservation of sensorimotor rhythms in mu(8-12Hz) and beta (13-30Hz) bands is crucial for discriminating different motor imagery tasks.

In seizure detection, our method achieves 91.9% sensitivity and 93.7% specificity, with an AUC of 0.947 (97.3% retention). This high performance is critical for clinical applications where missing seizures could have serious consequences. The superior preservation of both slow(delta/theta) and fast (beta/gamma) activities enables accurate detection of seizure onset patterns.

For emotion recognition, we maintain 97.7% of the original accuracy, demonstrating excellent preservation of affective patterns. The frontal alpha asymmetry and temporal dynamics crucial for emotion classification are well-preserved by our cross-channel attention mechanism.

5.4. Ablation Study

Table IV presents a comprehensive ablation study revealing the contribution of each component in our framework.

The frequency aware attention mechanism proves most critical, with its removal causing a 15.2% increase in NMSE. This validates our hypothesis that different frequency bands require specialized processing. Cross-channel attention contributes significantly (12.7% increase when removed), confirming the importance of modeling spatial correlations.

Replacing the Transformer with LSTM or CNN architectures results in substantial performance degradation (23.5% and 18.3% respectively), demonstrating the superiority of self-attention for capturing long-range dependencies. The multi-scale decoder is essential (17.1% increase with single-scale), as EEG signals contain patterns at multiple temporal scales.

The composite loss function proves valuable, with the MSE-only variant showing 14.3%worse performance. The frequency domain loss (9.8%contribution) ensures spectral fidelity, while SSIM loss (7.5% contribution) preserves morphological features.

TABLE III
DOWNSTREAM TASK PERFORMANCE AT CR = 8

Method	Motor Imagery (BCI)			Seizure Detection (CHB-MIT)			Emotion (DEAP)	
	Accuracy(%)	Retention	F1-Score	Sensitivity(%)	Specificity(%)	AUC	Accuracy(%)	Retention
Original Signal	82.4±2.3	100%	81.8±2.4	93.8±1.9	95.2±1.7	0.973±0.012	73.9±2.3	100%
OMP	73.5±2.7	89.2%	72.4±2.8	85.9±2.3	89.0±2.1	0.902±0.020	66.6±2.6	90.1%
AMP	74.0±2.6	89.8%	73.0±2.7	86.4±2.3	89.5±2.1	0.908±0.019	67.2±2.5	90.9%
ReconNet	75.6±2.6	91.7%	74.7±2.6	87.8±2.2	90.7±2.0	0.918±0.018	68.5±2.4	92.7%
ISTA-Net	76.2±2.5	92.5%	75.3±2.5	88.5±2.1	91.3±1.9	0.923±0.017	69.1±2.4	93.5%
DU-DORNet	77.4±2.5	93.9%	76.6±2.5	89.7±2.0	92.1±1.8	0.931±0.016	70.3±2.3	95.1%
CS-ConvRNN	78.3±2.4	95.0%	77.5±2.4	90.5±2.0	92.8±1.8	0.938±0.015	71.1±2.2	96.2%
TransformerCS-EEG	**79.7±2.3**	**96.7%**	**78.9±2.3**	**91.9±1.8**	**93.7±1.7**	**0.947±0.014**	**72.2±2.2**	**97.7%**

TABLE IV
ABLATION STUDY ON CHB-MIT DATASET AT CR = 8

Model Variant	NMSE	Change	SNR (dB)
TransformerCS-EEG (Full Model)	0.143±0.012	-	17.65±0.61
Transformer Architecture Components			
w/o Frequency-Aware Attention	0.165±0.015	+15.2%	16.34±0.68
w/o Cross-Channel Attention	0.161±0.014	+12.7%	16.47±0.67
w/o Position Encoding	0.158±0.014	+10.6%	16.67±0.66
Replace Transformer with LSTM	0.177±0.017	+23.5%	15.92±0.71
Replace Transformer with CNN	0.169±0.016	+18.3%	16.15±0.69
Sampling and Reconstruction Components			
w/o Learnable Sampling	0.159±0.014	+11.2%	16.58±0.67
w/o Channel Attention in Encoder	0.154±0.013	+7.8%	16.85±0.65
Replace Multi-scale with Single-scale Decoder	0.167±0.015	+17.1%	16.24±0.69
w/o Residual Connections in Decoder	0.156±0.014	+9.2%	16.75±0.66
Loss Function Components			
MSE Loss Only	0.163±0.015	+14.3%	16.42±0.68
w/o Frequency Domain Loss	0.157±0.014	+9.8%	16.72±0.66
w/o Structural Similarity Loss	0.154±0.013	+7.5%	16.87±0.65

5.5. Computational Analysis

Transformer CS-EEG achieves an effective balance between model complexity and performance. With 1.24M parameters, our model is larger than CS-ConvRNN(0.92M) but smaller than DU-DORNet (1.47M). The inference latency of 5.2ms for a 2-second segment (0.26% of signal duration) meets real-time processing requirements for clinical monitoring and BCI applications.

The memory footprint of 28MB for processing a 23-channel segment enables deployment on edge devices. The linear scaling with channel count and signal length (approximately 0.11ms per channel per second) ensures scalability to high-density EEG systems.

6. Conclusion

We presented Transformer CS-EEG, a novel Transformer based framework for compressed sensing of EEG signals that addresses the limitations of existing methods through innovative architectural designs. Our approach leverages self-attention mechanisms to capture long-range dependencies and complex spatial-temporal correlations, achieving 18% lower reconstruction error and 2.4dB higher SNR compared to state of-the-art methods at high compression ratios.

The key innovations—frequency-aware attention, cross channel modeling, and multi-scale reconstruction—prove essential for preserving neuro physiologically relevant information. Our framework maintains over 96% of task performance in motor imagery classification, seizure detection, and emotion recognition, demonstrating its practical utility for clinical and BCI applications.

With inference latency suitable for real-time processing and moderate computational requirements, Transformer CS-EEG offers a practical solution for efficient EEG acquisition in resource-constrained environments. Future work will explore adaptive compression strategies, incorporation of neurophysiological priors, and extension to multi-modal brain signal processing. The code and pre-trained models will be made publicly available to facilitate reproducibility and further research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This paper is supported by the Ministry of Education Industry-University Cooperative Education Project which is based on the Artificial Intelligence Teacher Training Project of the Industry-University Cooperation With the company of Zhongwu Technology (Project number: 230705160193805).

References

1. Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on information theory, 52*(4), 1289-1306.
2. Tropp, J. A., & Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory, 53*(12), 4655-4666.
3. Kulkarni, K., Lohit, S., Turaga, P., Kerviche, R., & Ashok, A. (2016). Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 449-458).
4. Zhang, J., & Ghanem, B. (2018). ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1828-1837).
5. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*(1), 5998-6008.